Domain General Rule Abstraction in 8-month-old Infants

A Senior Honors Thesis

Presented in Partial Fulfillment of the Requirements for graduation
with distinction in Linguistics in the undergraduate colleges
of The Ohio State University

by

Jason Grafmiller

The Ohio State University
June 2006

Project Advisor: Professor Peter Culicover, Department of Linguistics

Abstract

In language acquisition theory a crucial question centers on the degree of innate specialization for language learning. Over the past decade the importance of the ability to extract statistical information in both linguistic and non-linguistic domains has received considerable attention among linguists and cognitive scientists (Aslin, et al. 1998; Saffran, et al. 1996; Kirkham, et al. 2002). It is also well known that language acquisition must involve more than just extracting co-occurrence frequencies between items. Marcus et al. (1999) propose that there is also a mechanism designed to extract abstract, "algebraic" rules from linguistic data, though to date there has been no published studies examining this mechanism in non-linguistic domains. This study sought to replicate the findings of Marcus et al. with non-linguistic auditory and visual input. Results from three experiments show that 8-month-old infants are able to learn such rules from both linguistic and non-linguistic stimuli. This is taken as evidence that a rule abstraction mechanism of the kind proposed by Marcus et al. is part of the larger repertoire of domain-general learning mechanisms.

Since the middle part of the twentieth century, a major focus of research in generative grammar has been on developing theories of how a person acquires the language of his or her environment. From the very beginning, it was realized that this is a remarkably daunting task, and one that was inextricably linked with emerging theories of adult linguistic knowledge (Chomsky, 1965). Over the years, substantial experimental and naturalistic evidence has shown that children acquire language with extreme rapidity and accuracy, so much so that their levels of comprehension and production often appear to far exceed what might reasonably be expected of them given their limited linguistic experience (Pinker, 1994; Crain & Thorton, 1998; Guasti, 2002). Of particular importance is the remarkable lack of error and incorrect generalizations in the absence of any negative evidence. Learners seem to acquire rules that tell them what generalizations *not* to make, and it is argued that such rules require some kind of evidence, eg. negative feedback, which simply does not exist. These facts form the basis for the argument from the Poverty of the Stimulus, an argument that is the foundation for certain theories of language acquisition. Thus knowledge of the extraordinary complexity of language structure combined with this well-attested lack of negative evidence has led many linguists to the conclusion that certain aspects of linguistic knowledge are innate (inter alia Chomsky, 1965; Pinker, 1994; Guasti, 2002; Lidz & Gleitman, 2004b). Specifically, this knowledge is theorized to exist in the form of predetermined constraints on the variation of language structure. It is thought that such constraints prevent children from making many of the errors that are logically possible but are never seen in language development. In this view, innate knowledge is modeled as a set of universal parameters governing language structure. All languages can be characterized as unique sets of parametric values, and learning a language is simply a matter of adjusting the correct values for each parameter based on the information in the linguistic environment. The crucial point in this model is that the learner comes pre-equipped with a rich and highly structured knowledge of possible language variation, against which she must match what she hears around her to arrive at the correct grammar. Because of its innate nature, this knowledge has been called Universal Grammar or UG.

Support for this model comes from the supposed universality of certain aspects of linguistic knowledge. Again, these aspects are considered to be inviolable constraints on variation across all languages (Cook & Newson, 1996; Crain & Thornton, 1998; Crain & Lilo-Martin 1999; Guasti 2002). The core premise of this argument is that, given the high degree of variation within even a very narrowly defined linguistic community, it seems extremely unlikely that learners within a speech community should converge on the same grammar. The conclusion is that this convergence must be guided in part by some knowledge of the language's underlying structure. Such knowledge, it is claimed, could not come from the observable linguistic environment and so can only come from within the learner. That is, certain aspects of language structure must exist in the minds of all speakers from birth. On this view, UG knowledge is therefore part of our human genetic endowment. "Just as we cannot help but grow fingers and toes, because they are part of our biological blueprint for body development, we are preprogrammed to adhere to certain constraints on our linguistic development, according to the theory of UG" (Crain & Lilo-Martin 1999, p55). Additionally, UG knowledge is claimed to manifest itself at a consistent pace during learning across all languages. In other words, language acquisition tends to proceed through the same developmental stages and at roughly the same ages for all children and all languages (Pinker, 1994; Guasti 2002).

As the quote above suggests, analogies to other areas of human development have also reinforced the notion of a highly structured language acquisition device. Additionally, it has been assumed that this language acquisition device must be exclusive to the processing and acquisition of human language. "A theory that attributes possession of certain linguistic universals to a language-acquisition system…implies that only certain kinds of symbolic systems can be acquired and used as languages by this device. Others should be beyond its language-acquisition capacity" (Chomsky, 1965: 55). This makes strong claims about what can be considered a language, and more importantly what kinds of information can be learned by a human endowed with such a restricted language acquisition device. In essence, information that is structured like language but is not a language should not be learnable in this view. In principle, such a claim can be tested empirically; the question is how?

Critics of this theory contend that constraints on language structure might merely represent limitations on the ability to process highly complex information. Most of the information present in our environment is nowhere near the level of complexity inherent in natural language, and so it is understandable that any restrictions on higher-order processing might only become apparent in the close examination of language structure and acquisition. Advances in cognitive science have shown that infants and children do have surprisingly powerful domain-general mechanisms for processing statistical information, and it has been argued that such general mechanisms could be responsible for the acquisition of some rudimentary properties of language (Saffran et al., 1996; Aslin et al., 1998; Saffran, 2001; Kirkham et al., 2002; Creel et al., 2004). From this perspective, linguistic universals are hypothesized to arise from more general constraints on the processing of available information. In other words, learners converge on certain structures because 1) there are limits to the computational resources of the learner, e.g. memory storage, and 2) these limits act to restrict both the quantity and quality of data structures that can emerge from the processing of statistical properties in the linguistic environment. In addition, Maturational constraints on such general processing mechanisms could provide some explanation for the developmental trends seen in language acquisition (Newport, 1990; Hudson Kam & Newport, 2005).

It is important to point out that each of these approaches must make some concessions to the other (Gomez & Gerken, 2000). At some level, statistical information must be available to an inherited language acquisition device in order for it to correctly match the language of the learner's environment to the parameters of UG. In a more trivial sense, acquisition of a particular language's vocabulary and sound patterns certainly requires some manner of experience dependent learning. Likewise, it is fair to say that the human neurological architecture and its resultant restrictions on information processing mechanisms are themselves genetically predetermined in certain ways. Thus we are biologically predisposed to construct representations of any kind of data only in very specific ways. Essentially then, the difference between the nativist and empiricist camps lies in the degree to which cognitive learning mechanisms are considered to be specific to human language. The challenge for researchers becomes how to discern which language structures, if any, require some manner of innate, domain-specific learning device, and which structures can be built up from statistical information in the linguistic input.

As an initial point of attack, researchers have begun to look at some of the most basic properties of human language. One such property of language is the ability to make use of abstract representations such as arbitrary categories and algebra-like rules that assign variables to positions within a linear order. It is well known that all human languages make use of such rules and categories at various levels. At the phonological level, all languages make use of so-called phonotactic constraints, i.e. rules that determine the permissible ordering of phonetic segments within syllables and words. For example, speakers of English know that *brump* is a possible English word, while *rbupm* is not. At some point learners must recognize not only which sounds are permissible word-initially or word-finally, but also the precise linear order in which they can occur in these positions. Crucially, this type of rule can (must) extend beyond patterns based on perceptually unique items, where at the level of syntax, a language might have a rule requiring that a when a noun is modified by an adjective, the adjective must precede the noun it modifies. A learner must recognize that the only correct type of noun phrase is one that has this order. This point is discussed in more detail below.

In addition to understanding the kinds of structure inherent in natural language, it is also necessary to understand what types of learning mechanisms are available to infants in the very early stages of language acquisition. Experimental work suggests that to some degree infants are able to make use of statistical information present in their environment. Saffran, Aslin and Newport (1996) found that 8-month-old infants were sensitive to the transitional probabilities between syllables within a continuous stream of speech, and a similar experiment by Kirkham, Slemmer and Johnson (2002) found that this ability to exploit statistical information is not limited to linguistic or auditory input. The infants exhibited consistently longer looking times for the random sequences or sequences to which they were unfamiliar, indicating that they had learned the transitional probability between stimuli.

While the results from these experiments are indeed informative, evidence shows that simple co-occurrence probabilities do not explain all types of language phenomena. For instance, Johnson and Jusczyk (2001) conducted experiments that investigated the role of prosodic and co-articulatory cues in the word segmentation of infants. They contrasted the statistical and linguistic cues within speech streams and found that the statistical cues were overridden by more salient linguistic cues. In their experiments, infants preferred to use linguistic cues of stress and intonation rather than transitional probabilities to mark word boundaries in a word segmentation task. From this it appears that even in infancy, language learners are using more than just statistics to acquire language structure.

Marcus, Vijayan, Bandi and Vishton (1999), henceforth MVRV, claim to have found evidence for learning mechanisms capable of acquiring the kinds of algebraic rules described above. In their view, a rule stating that an English sentence can consist of a noun phrase followed by a verb phrase is a type of algebraic rule since it operates on abstract items such as noun phrases, which are only arbitrarily linked to their concrete realizations, e.g. "the boy, those feet", etc. MVRV contend that while earlier research has shown that infants are able to discriminate between familiar and novel sentences in artificial language experiments (Gomez & Gerkin 1997), this ability can be accounted for by statistical learning since the novel sentences were composed of the same words as the familiarization sentences—the infants could simply have learned the transitional probabilities between words.

In order to accurately test whether infants could learn rules, MVRV created three experiments using a simple artificial language comprised of 8 monosyllabic words. These words were arranged into 32 3-word sentences. Each sentence consisted of only two different words that were arranged into two different patterns: an *ABA* pattern and an *ABB* pattern. In the familiarization trials, infants heard sentences with either the *ABA* pattern or the *ABB* pattern, but not both. During the test trials infants were presented with sentences composed of entirely new words arranged into both the *ABA* and the *ABB* patterns. Using a basic head turning procedure, MVRV measured the amount of time the infants spent looking at the novel sentences. They found that infants consistently attended longer to sentences having a different pattern from the one they had been familiarized to. In addition, they ran a simulation of these experiments using a simple recurrent network (SRN) and were unsuccessful. They argue that the failure of their SRN was due to its inability to generalize to novel items. Because a network functions by altering the connection weights for individual words, it can only simulate rules by being trained on all items to which the rules apply. Therefore, MVRV concluded that the kinds of statistical mechanisms modeled by SRNs cannot account for how people generalize rules to novel items.

Researchers specializing in the design and implementation of SRNs have criticized the theoretical assumptions and conclusions as well as the experimental materials themselves (Seidenberg & Elman, 1999a; Seidenberg & Elman, 1999b). In general, the questions at hand are what defines a "rule", and how is a rule represented within the network architecture.[1] One objection was that the grammars did indeed contain statistical regularities despite what MVRV claim (Elmas, 1999; Negishi, 1999; Seidenberg & Elman, 1999a). In particular, infants "[were] exposed to a statistical regularity governing sequences of perceptually similar and different events" (Seidenberg & Elman, 1999a: 434). In other words, infants were not learning a true 'grammar', e.g. an algebraic *AAB* rule, but were merely incorporating the statistical relationships between similar and different items, e.g. same-same-different. Seidenberg and Elman are not clear however, on how this type of learning is different from the rule types of MVRV. Logically, a subject's recognition that a given string of items conforms to a 'grammatical rule', modeled here as *AAB*, entails that the subject recognizes that the first two items are identical and that both are different from the third. Whether we choose to call this

---

[1] Since the theoretical and mechanical issues surrounding the particulars of SRN design were not relevant to the focus of this study, they will not be addressed here.

learned pattern ‘same-same-different’ or ‘AAB’ is irrelevant. Both can be considered abstract patterns that can in principle be compared to the pattern of any novel input string. As some researchers have suggested, it is this very ability to recognize identities among elements in strings that underlies rule abstraction (Gomez & Gerken, 2000). I will return to this issue later.

The study of MVRV suggests that rule abstraction with linguistic data is present very early on, but to date there has been no published work examining such rule abstraction with nonlinguistic data. Importantly, MVRV make no claims about whether this ability is unique to language processing and leave open the question of whether it is part of a broader processing mechanism (note 7). It was the purpose of this study therefore, to examine the extent to which a rule abstraction mechanism of this type is active in other cognitive domains.

Experiment 1

Experiment 1 sought to first replicate the results of MVRV. The purpose of this was not only to add support to the results of the original study, but also to serve as a pilot test for the accuracy of the general procedure to be used later with different types of stimuli in Experiments 2 and 3.

Subjects

Subjects were sixteen 8-month-old infants, who were slightly older (1 month) than those who participated in the study of MVRV. Most of the earlier studies on statistical learning involved infants within this age group (Saffran et al. 1996; Aslin et al. 1998; Saffran, 2001; Kirkham et al. 2002), and there was no reason to suspect that this age difference would have any effect on the outcome of this study.

Stimuli

Stimuli for the training phase were created using 8 artificial monosyllabic words similar to those used by MVRV. All words had a simple consonant-vowel (CV) structure and were created with the text-to-speech synthesis program TextAloudMP3 using the AT&T Crystal voice. Several factors were taken into consideration during the construction of these words. First, limitations of the speech synthesis software precluded the use of particular CV sequences, and therefore it was not possible to duplicate the exact stimuli used by MVBV.[2] Second, as MVRV noted, it was necessary to ensure that there would be no chance that infants could be extracting statistical information based on sub-lexical features, and so the initial consonants were chosen to represent a range of articulatory features, e.g. voicing, aspiration, nasality, etc. This way there would be no consistent phonetic patterns like *voiced-unvoiced-voiced* or *stop-stop-fricative*. Third, in order to conform as much as possible to the stimuli of MVRVs' experiment, only two vowels were used, [ɑ] and [i], and these were each paired with one of four possible consonants, [b, t, l, f], creating a total of 8 possible words. Each word was also digitally adjusted for uniform duration and amplitude with the sound editing programs CoolEditPro and Praat. This was an additional control not carried out in the original

[2] MVRV provided the website address for the speech synthesizer used in their study, however the link was no longer active. An attempt to obtain the exact stimuli from the primary author was also unsuccessful.

Table 1. The 12 syllables used in the training and test phases (in IPA notation).

	Training	Test
A block	bɑ ti li fɑ	ko neɪ
B block	bi tɑ lɑ fi	peɪ ro

experiment.[3] Words were then assigned to one of two blocks, an A block or a B block, with each block consisting of four words (Table 1). The two blocks were: A block (*ba, ti, li, fa*), and B block (*bi, ta, la, fi*). Four test stimuli were created using the same method as the training words, only no phonetic segments present in the training stimuli were used in the test stimuli. Test words were also divided into two blocks, an A block (*ko, ne*), and a B block (*pe, ro*), and were adjusted to match the duration and intensity of the training words.

Procedure

Despite the desire to adhere as much as possible to their basic methodology, the procedural design of this experiment deviated significantly from those of MVRV in several ways. This was done for two reasons. One, a combination of resource limitations (time, lab space, subject availability, etc.) required that this procedure be much more efficient than the original. Rather than drawing out a replication of this study in a series of three experiments, as in the original, it was felt that the same results could be achieved through a single, more compact and carefully designed experiment, a more detailed description of which is given below. And two, subjects run on this experiment were also

[3] This was reported by Jon Slemmer in an informal presentation at the OSU Cognitive Development Center in the fall of 2005.

participants in a number of other studies taking place in the same lab. These other experiments were conducted using a preferential looking paradigm and reconfiguring the laboratory equipment between subjects and tasks would have placed considerable stress on both subjects and experimenters. It was felt that this experiment could easily be adapted from the full head-turn paradigm to the preferential looking paradigm without affecting the results, and furthermore, that successful replication using a different experimental paradigm would add robustness to the results. Thus the first significant deviation from MVRVs' study was the used of the preferential looking procedure.

Since this experiment was to be in effect three experiments collapsed into one, it was necessary to carefully design the experiment in such a way that it would account for all the possible confounds discussed by MVRV, such as the possibility that infants were merely habituating to patterns of sub-lexical features or else habituating to only patterns having two immediately reduplicated elements (p78). In the original study, these considerations ultimately necessitated the design of three different experiments in order to show sufficient evidence that infants were in fact learning the abstract rules as intended. Given that rule abstraction was successful in the original in all three cases, there was no reason to assume that a single experiment using randomized pattern conditions and carefully arranged stimuli could not achieve the same results with the same degree of reliability. Therefore, the second major difference in this experiment was that infants were randomly assigned to one of three possible training conditions (Table 2) and then randomly assigned to a test condition. This was unlike the study of MVRV where all infants were assigned to the same training and test conditions in each experiment.

Table 2. List of the six possible training-test conditions. Infants were each randomly assigned to one of these conditions.

Training	Test
ABA	AAB
ABA	ABB
AAB	ABA
AAB	ABB
ABB	ABA
ABB	AAB

In each training condition, stimuli consisted of a series of sentences composed of a subset of the 8 training words. Each sentence contained only two individual words, one from the A block and one from the B block, and these words were arranged according to the specified condition pattern, with a 250 ms pause inserted between each word. For example, the ABB condition would include all the possible training sentences where the second word was repeated, e.g. *ba ti ti*, *ta li li*, etc. All possible AB word pairs were used, giving a total of 16 sentences. Each sentence was repeated three times for a total of 48 training stimuli. Test sentences were constructed in similar fashion using the 4 test words. There were a total of four possible AB test pairs, thus there were only four test sentences for any given test condition.

Each infant was trained and tested individually, and sat on a parent's lap inside a small, enclosed room for the duration of the experiment. Inside the room, visual images were projected onto a screen positioned at a viewing distance such that the angle of the visual field was large enough that experimenters could clearly distinguish the direction of the subject's visual attention. Auditory stimuli were presented at a comfortable volume from speakers hidden behind the screen at the far right and left sides. Presentation of both auditory and visual items was controlled by a program written with the multimedia program Macromedia Director MX and run on a PC.

As described above, each infant was randomly assigned to one of the three patterns in the training procedure, and no subject was trained on more than one pattern. At the beginning of the training phase a blinking red circle was displayed to attract the infant's attention toward the lower center of a white screen. A short noise was also repeated in sync with the circle to help draw the infant's gaze to the screen. Once an experimenter had indicated the infant was looking at the center, the training sentences began playing from speakers on both sides of the booth. Each sentence was repeated three times, with a 1 second pause between sentences, and no sentence was repeated consecutively. The infants were presented with a total of 48 sentences and the entire familiarization process lasted little more than three minutes.

During the testing phase, each of the four test sentences was presented in a series of 12 test trials, where six trials consisted of sentences with the pattern familiar from training, and where the other six trials consisted of sentences with a novel pattern. Familiar and novel trials were presented in alternating order. At the beginning of each trial the same red circle flashed on the center of the screen to attract the infant's attention. Once the infant's attention had been directed to the center of the screen, an experimenter began the test trial. When the experimenter had indicated the infant was looking at the screen, a blue circle began flashing on one side of the screen. Once the infant was looking at the blue circle the experimenter indicated for the test sentences to begin playing.

In each trial, subjects were presented only the sentences with the familiar pattern or only the sentences with the novel pattern. Additionally, novel trials were presented on only one side of the screen while familiar trials were presented only on the other. For example, during testing an infant might hear the two ABB sentences "*ko ne ne, pe wo wo*" repeated from the speaker at the left side of the screen, and she might hear the two AAB sentences "*ko ko ne, pe pe wo*" repeated from the speaker on the right side. During each trial, test sentences were repeated with a 1.5 s pause inserted between repetitions until the infant looked away for 2 continuous seconds or 15 s had elapsed. To counterbalance for any preference of initial side or pattern, the specific side on which the familiar and novel trials were presented was chosen randomly by computer as was the consistency of the initial trial. The dependent variable was the total time an infant spent looking at the circles on either side of the screen. Looking times were recorded by an experimenter and verified later by off-line coding from videos of the sessions.

Results

All sixteen infants successfully completed 6 or more test trials. Overall, fourteen out of the sixteen infants showed a preference for the novel patterns during test, and a two-sample t test revealed significantly longer looking times for the novel pattern trials versus the familiar pattern trials ($t(25) = 2.2$, $p < 0.04$). Table 3 shows the mean looking times for familiar and novel patterns from infants in Experiment 1 along with results reported by MVRV from their first experiment.

Table 3. Mean looking time of familiar vs. novel sentences. For comparison, data reported by MVRV from their first experiment are provided as well.

	Mean looking time (s)		
	Familiar	Novel	
	3.147 (SE = 0.32)	4.141 (SE = 0.34)	$t(25) = 2.2, p < 0.04$
MVRV	6.3 (SE = 0.65)	8.3 (SE = 0.54)	ANOVA: $F(14) = 25.7, p < 0.001$

Figure 1. Mean looking times (in seconds) for familiar and novel trials from Exp 1. Figure 1b shows the mean looking times per trial for both novel and familiar patterns.

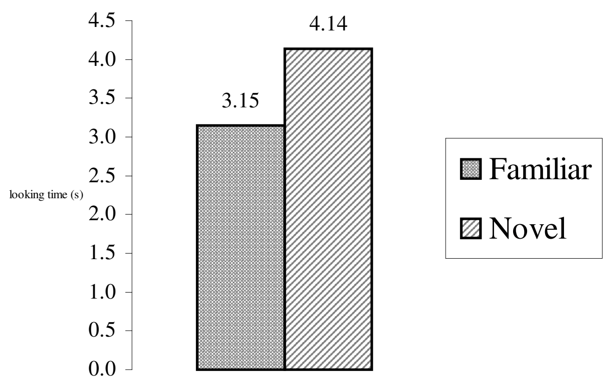
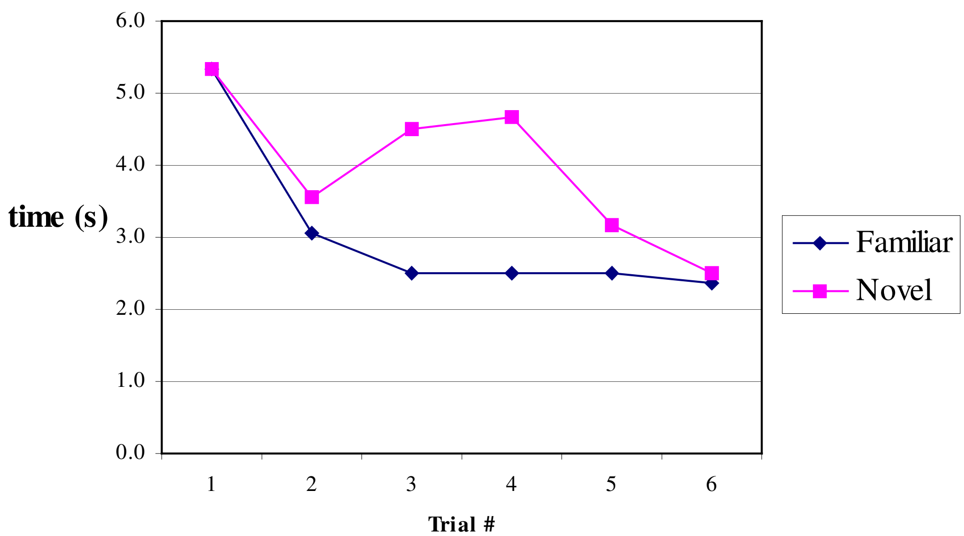


Figure 2. Exp. 1 mean looking times for novel and familiar patterns by trial.



Discussion

On the whole the experiment can be considered a successful replication of the findings of MVRV, however there were noticeable differences in the absolute looking times (nearly 50% shorter). There are several factors that may explain the difference in total looking times between these two studies. One explanation may be related to the difference in procedure. The original study was conducted using the head turn procedure developed for previous statistical learning studies (Saffran et al. 1996; Aslin et al. 1998), in which subjects' attention is directed to either side of a darkened booth by means of a flashing light. Although this study did not use such a procedure, employing instead a preferential looking procedure where subjects' gazes are directed to one side of a wide screen, there was no reason to believe that this difference in procedural paradigm should affect either the infants' ability to abstract rules or the validity of the experimenters' inferences of rule abstraction based on behavioral cues, i.e. looking times. The answer may be that the method for directing the infant's attention is simply more engaging in the head turn paradigm (a flashing light) than in the preferential looking paradigm (a blinking colored circle on a white screen). However, given the other differences between this experiment and those of MVRV, it is impossible to tell whether this explanation is correct, although intuitively it seems to make some sense.

Another explanation may be related to the quality of the stimuli themselves. Since we were not able to obtain the original stimuli for this experiment, or even the same program used to create the original stimuli, it is uncertain how accurately the original stimuli replicated natural human speech. The text-to-speech program used for this study, TextAloudMP3, was judged to be the most natural sounding out of several different programs, and in general the infants seemed to respond well to the stimuli. Nevertheless, some stimuli were distinctly artificial sounding and it is possible that infants would be more engaged with stimuli that more closely resembled familiar, natural English sounds. If the stimuli of MVRV were in fact superior, i.e. more natural sounding, then this fact might also provide some explanation for the differences in looking times.

Finally, infants who participated in this study also participated in other studies in the same lab. It is likely that subjects who had participated in tasks prior to this experiment were already fatigued, causing them to be less engaged during the procedure. This negative effect could easily be ameliorated in future studies.

Experiment 2

The successful rule abstraction in Experiment 1 provides additional support for the claim that infants are capable of learning simple algebraic patterns with linguistic stimuli at an early age. This is compatible with other studies that have shown that young infants can make use of statistical cues across syllables to segment words out of a steady stream of speech (Saffran et al. 1996; Aslin et al. 1998; Saffran 2001). Additional studies have provided evidence that this ability is not unique to language but is part of a more general ability to process transitional probabilities across any auditory stimuli. Saffran et al. (1999) familiarized 8-month-old infants to a steady stream of non-linguistic tones that were arranged into four three-tone words in a procedure identical to that used by Saffran et al. (1996) and Aslin et al. (1998) with the exception of the different stimuli. Infants were then presented words and part-words composed of tones during test. Results from this experiment were exactly in line with earlier results: infants showed a clear preference for the novel 'part-word' tone sequences. Saffran et al. (1999) thus concluded that the ability to segment sequences in a stream of sound based solely on the transitional probabilities between tokens must be part of a domain general learning mechanism, albeit a low-level one.

Based on a similar motive, Experiment 2 sought to investigate whether the particular learning mechanism responsible for abstracting the rules in Experiment 1 might also be underspecified for any particular type of input in the auditory modality. The goal was to show that infants could abstract the same rule pattern as in Experiment 1 with non-linguistic tone stimuli. If infants are indeed able to abstract such rules it was predicted that they should exhibit clear differential looking times for familiar versus novel test patterns. Additionally, it was initially assumed that infants would show greater looking times for novel patterns in the same fashion as Experiment 1.

Subjects

As in the first experiment, subjects in the tone experiment were sixteen 8-month-old infants. No infants who participated in the previous experiment also participated in this one.

Stimuli

In this experiment training stimuli consisted of sentences similar to those in experiment 1 except that instead of linguistic syllables, the "words" were synthesized tones. Eight different tones were generated in CoolEditPro and arranged into A and B blocks in the same fashion as the first experiment. The tones were constructed to vary along two parameters, pitch and modulation frequency. In order to eliminate the possibility that the infants were simply learning a pattern of change along any single parameter, the tone pairings were counter-balanced in such a way that some pairs would differ only in pitch, others would differ only in modulation and some would differ in both pitch and modulation. The duration and amplitude of these tones were consistent with the mean duration and amplitude of the word stimuli in the first experiment.

Test stimuli were created using the same software as the training stimuli. They consisted of four novel tones differing both in pitch and modulation frequencies from those of the training set. Again, two tones were assigned to the A test block and two were assigned to the B test block. As with the training stimuli, the tone pairs were counter-balanced to ensure variation across all possible combinations of parameters.

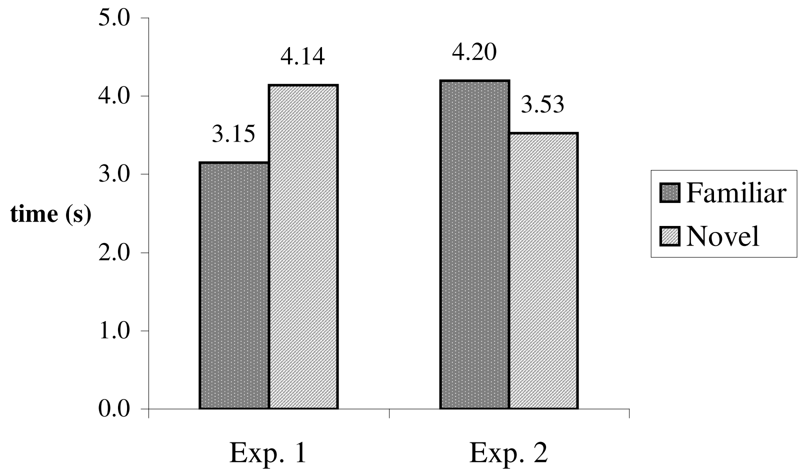
Procedure

With the exception of the difference in stimuli, the procedure for this experiment was exactly the same as in the first experiment.

Results

Fifteen infants completed 6 or more test trials. One subject failed to make it through 6 trials and so data from this subject was not considered in the analysis. A two-sample *t* test revealed no effect of test pattern on looking times for the non-linguistic tones ($t(27) = 0.72$, $p = 0.47$, ns) (Fig. 3). However, examination of mean looking times per test pattern within individual subjects suggested that there was a significant difference in the number of infants showing novelty preference versus the number of infants

Figure 3. Mean looking times for familiar and novel trials from Experiments 1 and 2. Analysis revealed that data from Exp. 1 was statistically significant ($t(25) = 2.2$, $p < 0.04$), however data from Exp. 2 was not ($t(27) = 0.72$, $p = 0.47$). This was primarily due to the differences in pattern preference.



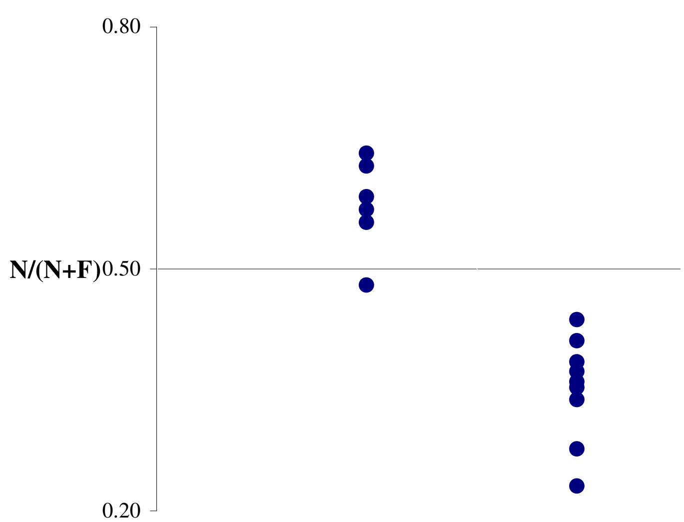
showing familiarity preference. To determine whether differences in preference were affecting the data, subjects were separated into two groups according to familiar versus novelty preference. Preference was determined using the ratio of mean looking time during novel trials (N) to the sum of the mean looking times for both novel and familiar trials ($N+F$). Ratios were then compared to the hypothetical mean 0.50 where novel and familiar looking times are equal. Subjects with a high ratio ($N/(N+F) > 0.55$) were considered to exhibit a novelty preference, i.e. they spent more time attending to the novel pattern, and subjects with a low ratio ($N/(N+F) < 0.45$) were considered to show a familiarity preference, i.e. they spent more time attending to the familiar pattern during test (Fig. 4). Six infants showed strong novelty preference ($M = 0.579$, $SE = 0.024$, $t(5) = 3.32$, $p < 0.02$), while eight infants showed strong familiarity preference ($M = 0.352$,

$SE = 0.021$, $t(7) = 6.95$, $p < 0.0001$). Only one subject (ratio = 0.48) had a ratio between 0.45 and 0.55.

Discussion

While these results are slightly at odds with the results obtained in the first experiment, they nevertheless suggest successful rule abstraction, since the objective of this experiment was to determine whether infants could clearly differentiate between the two types of patterns. Analysis of individual looking times showed that in general infants were able to distinguish the two test patterns, although they exhibited different preferential looking patterns. Given that infants were not subjected to a full habituation procedure, which involves presenting infants with the stimuli until their looking time has dropped below a predetermined threshold, this kind of result is not entirely unexpected.

Figure 4. The *N*/(*N*+*F*) ratios of subjects showing novelty preference (left) and familiarity preference (right).



Full habituation ensures that infants have become bored with the stimuli, resulting in a strong preference for novel stimuli during test. With the design of this experiment however, there is no guarantee that infants have become bored with the familiar pattern. Basically, some infants in this experiment simply may not have tired of the familiar pattern sufficiently to cause them to disprefer it when presented it at test.

Why there would be a difference between this experiment and the first is not entirely clear, but there may be an explanation based on the infants' pre-exposure to the linguistic stimuli used in the first experiment. Eight-month-old infants have had considerable exposure to their native language, and so it is possible that they have already been developing the necessary apparati for processing such linguistic data. Seidenberg et al. (1999) bring up this point when disputing MVRVs' claim that SRNs cannot replicate their results, and it is no less valid here. The fact that infants have "already developed a rich representation of the structure of acoustic and speech events on the basis of several thousand hours of exposure to examples" (Seidenberg &Elman, 1999a: 435) is significant since it is unlikely that infants have had a degree of exposure to the types of artificial tones used in the second experiment equivalent to what they have likely had for language sounds. In other words, infants are just not used to processing these particular kinds of sounds, and so they are slower to recognize patterns between them.

At first glance, this explanation seems problematic when considering the results obtained by Saffran et al. (1999). In their experiment with tone sequence segmentation (Exp. 3) they found that their infant subjects showed a clear preference for the part-word sequences, i.e. the novel forms (part-word M = 6.92, word M = 5.88) (p45). This seems to suggest that infants do not have problems processing unfamiliar stimuli, however I think there are a couple of reasons why this case is different.

First, the number of both stimulus tokens and stimulus types presented during familiarization was greater in the experiment of Saffran et al. In their experiment, infants were familiarized to 180 instances of non-linguistic tones, selected from a set of 12 different possible tone types. In the present experiment, infants heard only 144 tokens, taken from a set of only 8 possible tone types. While these differences may turn out to be irrelevant, it nevertheless seems plausible that the greater variety in tone types and the larger number of tokens (an increase of 25%) in the experiment of Saffran et al. were enough for infants to acquire the necessary processing familiarilty to fully familiarize to the appropriate sequences.

The other reason that we see a difference between these two experiments is that they are testing potentially very different abilities. The study of Saffran et al. (1999) was designed to investigate the ability of 8-month-olds to learn the transitional probabilities between adjacent tones in a steady stream, and use those statistics to discriminate between particular sequences of tones. Alternatively, the experiments created by MVRV were designed to provide evidence that infants are capable of learning rules over abstract variables, rules equivalent to something like “item X is the same as item Y”. Importantly, MVRV argue that this rule learning mechanism is quite different from the one responsible for learning simple statistics. If this is correct then there is no reason assume that learning with different types of stimuli should proceed at the same pace in both mechanisms.

Related to this is the possibility that these mechanisms become active at different points in an infant's development. On the assumption that learning the transitional probabilities across adjacent items is a less demanding task than extracting abstract rules, it is reasonable to expect that such a statistical learning ability might be present at an earlier age. While most statistical learning studies have involved infants 8 months or older, Kirkham et al. (2002) investigated statistical learning in the visual domain with infants at the ages 2, 5 and 8 months, and found that all age groups successfully discriminated between familiar and novel sequences. These results suggest that statistical learning mechanisms are active at least as early as 2 months. To date the youngest infants studied on rule extracting tasks have been the 7-month-old infants who participated in the study of MVRV.

Experiment 3

Experiments 1 and 2 demonstrated that infants are able to learn abstract algebraic rules instantiated with both linguistic and non-linguistic stimuli in the auditory modality. The goal of Experiment 3 was to answer the question of whether such rules can be learned from stimuli presented in the visual modality. This experiment is currently in progress.

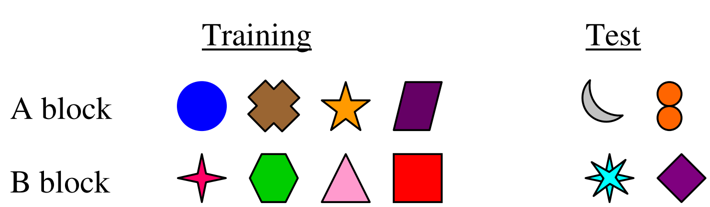
Subjects

Following the same procedure as Experiments 1 & 2, sixteen 8-month-old infants will participate in this experiment. Again, no subjects who participated in either of the previous experiments will participate in this one.

Stimuli

The stimuli used for this experiment consisted of 12 colored shapes. Eight shapes were used for the familiarization phase and the remaining four were used only during the test phase. Each shape stimulus was created as a very short movie (0.66s) using the program Macromedia Flash MX Professional. Each stimulus movie was designed to mimic a looming effect where the shape would appear to grow from a height of # to $. The intention was that this effect would help to maintain the infants' attention during the procedure (Kirkham et al. 2002). The training shapes were grouped into two blocks of four each, an A block (*blue circle, brown X, orange star, indigo parallelogram*) and a B block (*fuscia cross, green hexagon, pink triangle, red square*). Similarly, the test shapes were arranged according to the 2 x 2 pattern used in Experiments 1 and 2: an A block (*gray moon, tangerine snowman*), and a B block (*purple diamond, turquoise sun*).

Table 4. The 12 shape stimuli used in Exp. 3



Procedure

The procedure for this experiment was identical to the first two with the obvious exception that there were no sounds presented during trials. Additionally, the familiarization trial was slightly longer (approx. 4 min) due to the added length of the visual stimuli.

Results

This experiment is currently still in progress. Presently there is too little data to draw any significant conclusions.

General Discussion

The objective of this study was to investigate the extent to which human infants are able to extract certain kinds of patterns from a relatively small amount of input. Specifically, these patterns are not simply statistical relationships between individual tokens of the input, but must be represented at some level as specific linear ordering relationships between two or more abstract variables or placeholders. Such an abstract representation is necessary for infants to generalize the pattern learned during training to the test input. The question at hand was whether this rule abstraction ability is active both linguistic and non-linguistic domains.

Experiment 1 was a successful replication of the study conducted by Marcus et al. (1999) that demonstrated rule learning in the linguistic domain with 7-month-old infants. In this experiment, 8-month-old infants were familiarized to one of three possible 3-word patterns, *AAB, ABA,* or *ABB*, where the words consisted of artificial English syllables. During testing, infants were presented both the familiar pattern and a novel pattern, both of which were constructed using entirely novel syllables. In direct correspondence with the results of MVRV, infants in this experiment showed clear preference for the novel patterns during test, suggesting that they had successfully learned the intended rule. Minimally, this provides further evidence that infants are able to acquire abstract rules based on very small amounts of linguistic input.

In Experiment 2 and 3, I sought to determine whether such rules could also be learned from non-linguistic auditory and visual input. The results of Experiment 2 demonstrated that these kinds of rules could indeed be learned by infants, as determined by their ability to differentiate between a familiar and a novel test pattern, although here subjects did not show a consistent preference for either pattern. It was suggested that this result was due to greater unfamiliarity with the auditory tones which resulted in a greater processing load on the subjects. This increased processing load caused some infants to fail to fully habituate to the training pattern, thus leading them to prefer the familiar pattern during testing. A related possibility could also be that these types of non-linguistic tones are naturally more difficult to process, although at the moment I have no information as to why that should be the case.

Even though there are no results yet from Experiment 3, it is worthwhile to take a moment to speculate on the possible results and their implications when considered with those from Experiments 1 and 2. In a way, I see the successful learning of the rule in Exp. 3 as the less interesting of the two possibilities, despite being the intended goal of this study. Success here would suggest that we are simply seeing a truly general cognitive mechanism, i.e. an ability that can learn patterns from any type of stimulus. Such a mechanism would be exactly in line with various other mechanisms that have been shown to be independent of cognitive domain, e.g. those responsible for learning transitional probabilities. On the other hand, failure to learn the rule with visual stimuli would reflect successful rule abstraction in the auditory domain and a failure to apply such rules to visual input. This would suggest that the mechanisms involved in the rule abstraction process are especially (or exclusively) sensitive to auditory information during infancy. This would be especially interesting given that adults are able to learn artificial grammars with both auditory and visual stimuli (Reber, 1976; Altmann et al., 1995). To examine this, more complex experiments could be run with children and adults at various stages of linguistic development. Additionally, further experimentation with visual stimuli (linguistic and nonlinguistic) might reveal insights about possible constraints/limitations on learning in the visual domain. Given that sign language acquisition proceeds at roughly the same pace as spoken language acquisition, it would seem that any strong auditory bias would be either active only very early in development or only trivially relevant for language acquisition.

Speculation aside, what do these results tell us about the nature of the mechanisms involved in early language learning? Admittedly, the implications of this study are unclear. The kinds of constraints and parameters considered by nativists to require UG knowledge, e.g. binding relationships on antecedents and pro-forms or whether a language can lack an overt subject, are far removed from the simple, perceptually-based patterns learned in these experiments. One possibility is that a mechanism capable of extracting this kind of information is but one more preliminary statistical processing mechanism necessary for parameter setting in UG, similar to the ability to compute transitional probabilities. For example, a very simplistic characterization of transitive sentences involves the specific order of the subject and object noun phrases relative to the verb. Thus an English learner must recognize that the

English sentences have a SVO pattern, while a Japanese learner must recognize that Japanese sentences have a SOV pattern. These recognitions might then feed into more complex representations via some kind of bracketing or hierarchical structure building device, something like [S[VO]] or [S[OV]]. This in turn would give a learner ample evidence for setting a head-direction parameter to either head-initial or head-final (Guasti 2002).

An issue directly related to this concerns the types of output generated by this rule learning. MVRV simply refer to these as algebraic rules ranging over abstract variables, however they overlook the fact that the variables represented by such rules are in fact perceptually bound. That is, the variables are not truly abstract in the more meaningful ways in which linguistic structures such as syntactic categories are (Gomez & Gerken, 2000). A rule of the type *ABA* is different from one of the type *noun-verb-noun* because we know that the two nouns are not identical in form, e.g. *Pat likes Sandy*, but are the same in that they are both members of a larger class. It is this level of abstraction that is necessary for learning the kinds of syntactic structures that occur in natural language, and generalizing these structures to novel productions. The question for empiricist theories is how this perception-based pattern could be in effect 'ramped up' to the higher level of category-based patterns. It is not clear how this might occur, although intuitively all it seems to require is that this rule learning can function on any type of linguistic item. Perhaps, having learned the pattern based on perceptually similar language items, the learner might view this pattern as applying to any type of linguistic objects. Looking at statistical learning with language sounds and tones, Saffran (2001) found that domain-general implicit abilities can generate domain-specific knowledge, claiming that the output of statistical learning is determined by the input. If categories (syntactic or otherwise) can be derived statistically from distributional evidence in the linguistic input, these categories might naturally be treated as linguistic objects, allowing them to be used as input to any other learning mechanisms. Traditionally, syntactic categorization has been considered to be problematic for learning-based theories of acquisition (Pinker, 1987; 1994), but recent studies have shown that there may be sufficient distributional evidence at least for the elementary categories noun and verb (Cartwright & Brent, 1997; Mintz et al, 2002; Mintz, 2003; Hohle et al, 2004). More research will certainly be necessary to determine if this is indeed a possibility.

In summary, these experiments showed that the ability to extract very simple algebra-like rules is mostly likely not limited to the domain of language, but one more of the general cognitive tools human beings use to make sense of their environment. The extent to which these results help to tease apart the degree of innateness necessary for language acquisition remains uncertain however. But to some degree this is to be expected. The goal of this study was certainly not to provide a definitive answer to the question of the cognitive exclusivity of language, rather it is hoped that it will serve as a point of departure for more extensive research in the future.

References

Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*(4), 899-912.

Aslin, R. N., & Fiser, J. (2005). Methodological challenges for understanding cognititve development in infants. *Trends in Cognitive Sciences, 9*(3), 92-98.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-olds. *Psychological Science, 9*(4), 321-324.

Buchner, A. (1994). Indirect effects of synthetic grammar learning in an identification task. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*(3), 550-566.

Cartwright, T. A., & Brent, M. R. (1997). Syntactic categorization in early language acquisition: formalizing the role of distributional analysis. *Cognition, 63*, 121-170.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Cook, V., & Newson, M. (1996). *Chomsky's Universal Grammar*. Cambridge, MA: Blackwell Publishing.

Crain, S., & Lillo-Martin, D. (1999). *An Introduction to Linguistic Theory and Language Acquisition*. Malden, MA: Blackwell Publishing.

Crain, S., & Thornton, R. (1998). *Investigations in Universal Grammar*. Cambridge, MA: MIT Press.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*(5), 1119-1130.

Dominey, P. F., & Dodane, C. (2004). Indeterminacy in language acquisition: the role of child directed speech and joint attention. *Journal of Neurolinguistics, 17*, 121-145.

Elmas, P. (1999). Do infants learn grammar with algebra or statistics? *Science, 284*, 435-436.

Gomez, R. L. (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology, 33*, 154-207.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109-135.

Gomez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*(5), 178-186.

Gomez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science, 7*(5), 567-580.

Guasti, M. T. (2002). *Language Acquisition: The Growth of Grammar*. Cambridge, MA: MIT Press.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? . *Science, 298*, 1569-1579.

Holhe, B., Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M. (2004). Functional elements in infants' speech processing: the role of determiners in the syntactic categorization of lexical elements. *Infancy, 5*(3), 341-353.

Hudson-Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: the roles of adult and child learners in language formation and change. *Language Learning and Development, 1*(2), 151-195.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition, 83*, B35-B42.

Lidz, J., Gleitman, H., & Gleitman, L. (2003). Understanding how input matters: verb learning and the footprint of Universal Grammar. *Cognition, 87*, 151-178.

Lidz, J., & Gleitman, L. (2004a). Argument structure and the child's contribution to language learning. *Trends in Cognitive Sciences, 8*(4), 157-161.

Lidz, J., & Gleitman, L. (2004b). Yes, we still need Universal Grammar. *Cognition, 94*, 85-93.

Marcus, G. F. (1999). Do infants learn grammar with algebra or statistics? *Science, 284*, 436-437.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*, 77-80.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*, 91-117.

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The Distributional structure of grammatical categories in speech to young children. *Cognitive Science, 26*, 393-424.

Negishi, M. (1999). Do infants learn grammar with algebra or statistics? *Science, 284*, 435.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14*, 11-28.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127-162.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General, 119*(3), 264-275.

Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition*. Hilsdale, NJ: Lawrence Erlbaum Associates.

Pinker, S. (1994). *The Language Instinct*. New York: William Morrow Company.

Pinker, S., & Jackendoff, R. (In press). The faculty of language: what's special about it? *Cognition*, 1-36

Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review, 19*, 9-50.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Experimental Psychology, 6*, 317-327.

Reber, A. S. (1976). Implicit learning of synthetic languages: the role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 88-94.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*(3), 219-235.

Saffran, J. R. (2001). Words in a sea of sounds: the output of infant statistical learning. *Cognition, 81*, 149-169.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27-52.

Seidenberg, M., & Elman, J. (1999a). Do infants learn grammar with algebra or statistics? *Science, 284*, 434-435.

Seidenberg, M., & Elman, J. (1999b). Networks are not 'hidden rules'. *Trends in Cognitive Sciences, 3*, 288.

Seidenberg, M., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science, 23*(4), 569-588.

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences, 4*(4), 156-163.

Zizak, D. M., & Reber, A. S. (2004). Implicit preferences: the role(s) of familiarity in the structural mere exposure effect. *Consciousness and Cognition, 13*, 336-362.